Fast Sub-pixel Precision Variable Block Size Motion Estimation

Ying Zhang+*, Wan-Chi Siu* and Tingzhi Shen+*

*Depart of Electronic and Information Engineering
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong, China

+Department of Electronic Engineering
Beijing Institute of Technology
Beijing, 100081, China

Abstract— **Fast sub-pixel precision variable block size motion estimation is a key issue for real-time application of the H.264. Many fast schemes have been proposed for fast variable block size sub-pixel motion estimation. Most of them are based on the analysis of the current frame. In this paper, we propose a new way to classify macroblocks of the current frame making use of the statistical information of the classified block types and motion activities of the previous frame, and the motion activities of the current frame. This forms a new and comprehensive scheme for fast variable block size motion estimation with sub-pixel refinement. With this prior knowledge, we can make decision on early rejection of sub-partition modes and early termination to skip candidate checking points. The new algorithm is able to alleviate substantially the computational effort of sub-pixel motion estimation. Extensive experimental work has been done, results of which show that our approach gives a speedup of 5 times over that of the fast algorithm in JM and a similar peak signal-to-noise ratio (PSNR) as the full search.**

Index Terms **–** Video Coding, Motion estimation, Motion Vector, Variable Block Size, and Fast Block Type Decision

1. INTRODUCTION

H.264 is an advanced video coding standard. Its high performance in coding is a consequence of the requirement for high computational complexity of this standard. Its motion estimation strategy supports 7 modes, which stand for 16x16 block size and its sub-partitions. Many fast motion estimation algorithms in the literature are designed for the traditional motion estimation strategy which supports only 16x16 block size[1-7] or does not have sub-pixel motion estimation[8]. An efficient sub-pixel precision variable block size motion estimation algorithm, which can accelerate the complex processing of mode selection and reduce the heavy burden of sub-pixel motion estimation is very desirable.

We can make an early mode decision after checking a small set of modes. A threshold is required to compare with the SAD for decision. The thresholds of most previous algorithms are based on either the analysis of the current frame or the statistics of amount sequences [9-10]. It is desirable that the threshold can reflect the character of a sequence and it can also adaptive to the change between frames. Only using a single threshold for decision cannot guarantee that the current mode and the temporarily best search point are good enough. The risk to skip other modes and checking points could be rather high. It is found that block classification (which implies whether or not the block belongs to the same moving object with its neighboring blocks) with a directional search strategy[11] can provide a noticeable improvement in the traditional motion estimation algorithm. In this paper, we analyze the motion information in the H.264 encoder, then apply them to help us make early mode decision on the block type before we finish all the block sizes. It is also used to assist the judgment of early termination in both integer-pixel motion search stage and sub-pixel refinement stage. The new algorithm can achieve an average speed-up of about 5 times of the search time compared to that of the original fast algorithm for the tested sequences using all the block sizes, while its PSNR and bits performance are very similar to the full search approach.

2. ALGORITHM DEVELOPMENT

2.1 Mode prediction based on the relationship of the motion estimation results of different block sizes

2.1.1 Relationship between results of temporal best mode and final best mode of mode decision

The search results of using different block sizes are inter-related. In our algorithm, we start to search with a block size of 16x16 in the first stage, and then check 16x8 and 8x16 block sizes in the second stage. The 8x8 block size and other partitions are considered in the third stage. The decision mode obtained from the first two stages is entitled as the “temporal best mode”. If P16x16 is the temporal best mode, it means that a big block size is selected among the three modes. This indicates that the moving object is probably not smaller than P16x16 and the chance to choose a mode even smaller than P16x8 and P8x16 is usually low. Sample statistics on this point are shown in table 1. These are the results of 18 test sequences which include Claire, Akiyo, Mother & daughter, Hall, Silent, Highway, Container, Erik, Paris, Foreman, Football, Waterfall, Coastguard, Bus, Tempete, Stefan, Flower and Mobile. For space limitation, we show only the result of three typical sequences and the average result of 18 sequences in table1.

Table 1. The chance of choosing 8x8 mode if one of the P16x16, P16x8 or P8x16 modes is selected

Sequences	P16x8 or P8x16 is selected and P8x8 is the best one among all inter modes		P16x16 is selected and P8x8 is the best one among all inter modes	
claire	49824	99%	325	1%
foreman	28342	96%	1329	4%
stefan	26851	88%	3526	12%
Avg. of 18 sequences	31710	94%	1909	6%

2.1.2 Mode prediction

The temporal best mode is used to predict the final best mode in our algorithm. If the temporal best mode is a mode with big block size and the SAD value for the best match in this mode is small enough, we then consider skipping other modes (8x8 and its sub partition). We suggest further using block classification as a means to achieve this.

2.2 The assistant motion information

2.2.1 Block classification according to the variable block sizes motion information

The block classification is designed to imply the motion characteristic of a block. During the motion estimation procedure, the motion vectors of the neighboring macro blocks are used to predict the motion vector of the current macro block before the pattern search stage. If the final motion vector obtained after the pattern search is not the same as the predicted motion vector. This kind of MBs are classified into the special MB group (Group S), whereas the other MBs are classified into ordinary macroblock group (Group O). Among the predicted motion vectors and the final motion vector of the 7 partitions, only those of the best mode are used in the block classification process. The special MBs and the ordinary MBs of the current frame can be "prejudged" according to the block classification of the previous frame. The method is to set the special macroblock and its 8 surrounding MBs as prejudged special MBs, and set the other MBs as prejudged ordinary macroblocks.

2.2.2 Distribution of SAD value of the best mode in successive frames

The SAD statistics of the successive frames are similar to each other. The SAD of the previous frame can also be used to predict the SAD statistics of the current frame. In fig 1, "SAD" value represents the SAD values of the best mode of blocks in frame 3 to frame 5 of the foreman sequence. Since most of the SAD values of these frames are distributed within range from *0* to *2000*, only SAD values in this range are shown (in the horizontal axis) in fig.1 for illustration. These values are divided into 16 groups. The numbers of blocks which have SAD values larger than the minimum values of respective groups are indicated in the vertical axis.

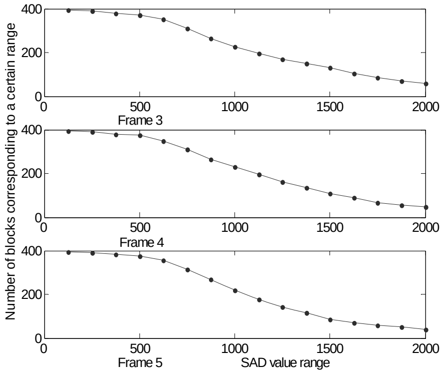

Figure 1. No. of blocks corresponding to a certain range of SAD

Hence, we can find a threshold by looking up the SAD value corresponding to the block number which is equal to the number of ordinary blocks in the previous frame in the "SAD value-block number distribution" graph (fig.1) of the current frame. That is, we choose thresholds in a way which makes use of the number of ordinary blocks of successive frames.

An upper bound and a lower bound are used in order to avoid threshold obtained from the actual data becoming too large or too small. The upper bound and lower bound are set to 800 and 1500 according to the results of a large set of experiences.

2.3 Sub-pixel motion estimation elimination

In our analysis, we further confirm that mode decision does not depend on the sub-pixel motion estimation. We can consider sub-pixel motion estimation after the mode decision. This means that we can perform integer-pixel motion estimation for each mode, choose the best mode according to their rate distortion values, and then perform sub-pixel motion estimation for only the best mode from the point given by the motion vector of the best integer mode.

2.4 Overall structure of algorithm

The flow chart as shown in fig.2 combines our fast algorithm with the mode decision procedure of the H.264 software (JM12.2). There are 4 modules in this approach, which are highlighted in bold. Note that the integer-pixel motion estimation stage implies also variable block size motion estimation, and it has 3 stages. There is always one case leading to a conditional execution of "early termination" in stage1, stage2 and stage3, respectively. In each of these three stages, the test point of the condition is positioned after checking all predictors, which are motion vectors of the neighboring blocks in the spatial and temporal domain. The conditions are:

1) The current MB is prejudged as an ordinary block.

2) The current minimum SAD cost for the block is smaller than the threshold for early termination, which is calculated at the end of the motion estimation process of the last frame.

3) The current minimum SAD cost for the block is smaller than the minimum SAD cost of the co-located MB.

If the conditions are true, an early termination will be executed, which means the best predicted motion vector (PMV) will be selected as the final motion vector (MV) of the current MB of a certain mode. If the MV of P16x16 is good enough, all others modes will be skipped.

There is another case of fast mode decision: Early Mode decision2, which can be found at the end of stage2. The condition for Early Mode decision2 is that the best mode among the above three modes (P16x16, P16x8 and P8x16) is P16x16.

After all inter and intra modes are checked, the best mode will be selected in the mode decision module. Then the statistical motion information of the MB is calculated and stored according to the best mode. If the PMV of the current MB is equal to its final MV in the best mode, the number of ordinary blocks will increase. The SAD values of the best modes of the current MB are stored to form the SAD values the macroblocks in the frame.

Sub-pixel motion estimation is be done on the best mode only. The directional information is used to accelerate the sub-pixel motion estimation.

There is also a chance for early termination after half-pixel motion estimation. The conditions are:

1) The current MB is prejudged as an ordinary block.

2) The motion vector of the co-located MB is zero.

3) The motion vector of the current MB with Half-pixel refinement is zero.

If all the MBs have been processed, the statistical motion information of the frame is than calculated, and a threshold for early termination will be generated from the result of the statistics. The SAD values of all MBs (16x16) are reordered in the previous process, when the statistics of each MB is calculated. The range of these SADs, from the minimum value to the maximum value, is divided into 16 groups and the SAD values will be classified into these 16 groups according to their values. Then, a new distribution will be obtained from the divided-SAD value ranges and the numbers of blocks, which have a SAD value larger than the minimum value of the ranges, are shown in fig.1. From this new distribution, we can find the threshold for early mode decision by looking up the SAD value corresponding to the block number equals the number of ordinary blocks of the current frame, which is stored as the statistics to be used in the next frame.

Note that in integer-pixel search part, a diamond shaped search pattern is applied to search around the predictor, and the directional information introduced in [10] is considered to reject some of the checking points. In sub-pixel refinement part, both the half-pixel refinement and the quarter-pixel refinement also use the directional information so that only 6 of the 8 half-pixel points (or quarter-pixel point) are going to be checked.

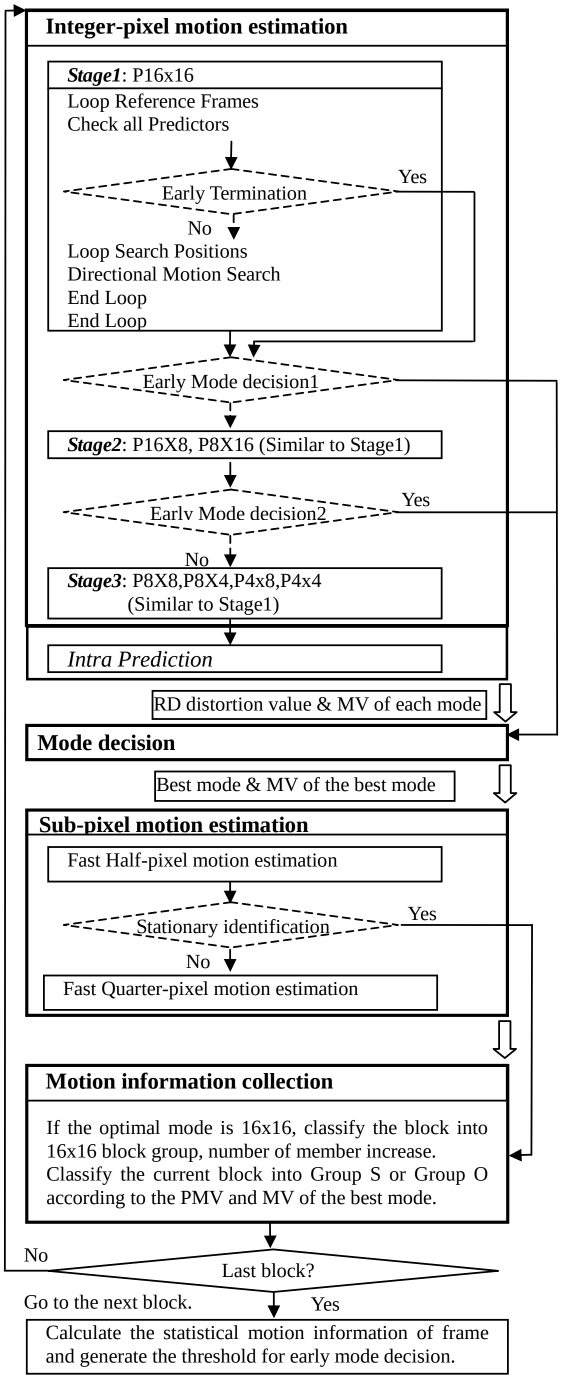

Figure 2. Flow chart of the suggested algorithm

3. EXPERIMENTAL RESULTS

The proposed fast variable block size motion estimation algorithm is examined by comparing the full search (FS) motion estimation algorithm, and the Enhance Predictive Zonal Search (EPZS) algorithm of JM12.2 of the H.264.

The PSNR listed in the Table 2 is the average PSNR value of the luminance component of the decoded frames. "Bits per pixel" gives the average value of the bits used to code one pixel, while the frame number to be coded is set to 150 for all the sequence. "SpUp" refers to speed-up, which is the ratio of computational times between the full search and other fast search algorithms tested. The software was the JM12.2 encoder. We used an IBM Notebook with Inter(R) Core (TM) 2 Duo CPU T7300 2.0G and 2GB RAM. A large number of sequences with the CIF format were tested. The results of the 18 sequences are arranged from low, medium to high motion activities as shown in table 2. The experimental work was conducted with a constant QP, which was 28. The search range was 32, and the RDOptimization option was set to 1, which is the high complexity mode.

Table 2. Performance comparison of motion estimation algorithms.

Sequences	**FS**			**EPZS**			**PRES**		
	PSNR	Bits per pixel	SpUp	PSNR	Bits per pixel	SpUp	PSNR	Bits per pixel	SpUp
claire	41.75	0.0287	1	41.72	0.0285	7	41.67	0.0302	38
akiyo	39.86	0.0307	1	39.85	0.0307	7	39.77	0.0332	41
mother_ daughter	38.98	0.0505	1	38.99	0.0505	9	38.95	0.0515	43
hall	37.94	0.0914	1	37.92	0.0912	9	37.83	0.0809	53
silent	36.15	0.0854	1	36.15	0.0858	10	36.16	0.0800	60
highway	38.62	0.0885	1	38.60	0.0907	10	38.65	0.0983	52
container	36.10	0.0703	1	36.08	0.0705	11	36.02	0.0772	61
erik	37.15	0.0759	1	37.13	0.0757	10	37.21	0.0694	55
paris	35.99	0.2054	1	35.99	0.2054	10	35.94	0.1679	63
foreman	36.98	0.1399	1	36.96	0.1414	10	37.02	0.1423	53
football	37.76	0.2253	1	37.74	0.2280	13	37.83	0.2262	67
waterfall	35.03	0.1112	1	35.04	0.1118	13	35.05	0.1149	64
coastguard	35.30	0.4836	1	35.30	0.4840	15	35.29	0.4649	76
bus	35.61	0.467	1	35.63	0.4651	14	35.66	0.4461	68
tempete	35.47	0.5049	1	35.46	0.5049	14	35.43	0.4832	68
stefan	36.33	0.4596	1	36.32	0.4606	13	36.34	0.4360	63
flower	36.09	0.6371	1	36.08	0.6376	12	36.20	0.5931	63
mobile	35.12	0.8199	1	35.11	0.8157	15	35.10	0.7248	78
Average	**37.01**	**0.2542**	**1**	**37.00**	**0.2543**	**11**	**37.01**	**0.2400**	**60**

From Table 2, the EPZS algorithm in JM12.2 is faster than Full Search, while the PSNR is 0.01dB lower. Compare with the EPZS motion estimation algorithm, our proposed algorithm is more accurate and faster (5 times faster) than EPZS, while its PSNR and bit rate is very similar to that of the full search algorithm. The proposed algorithm can provide more consistent motion vector field so that the proposed method can even achieve lower Bits per pixel while maintaining same PSNR on average. Furthermore we have also done experimental comparisons among QP values equal to 26-34. The results show that our algorithm is more efficient for all tested QPs.

4. CONCLUSION

We have done an analysis on the statistics of mode selection of variable block size motion estimation. We found that there is more than 80% of the computation does not provide a better result (QP=28). We have also done much work on an analysis of the relationship between the motion estimation result of different block sizes and the motion information of successive frames, which leads us to find a better chance to make early termination with little cost on the search result. The proposed new method for early mode decision of variable block size motion estimation relies heavily on the statistical information of the motion characteristics of the successive frames. Results of our experimental work show that our approach has a speed-up of 60 times compared with the original sub-pixel precision variable block size full search motion estimation on average. Besides what we have achieved, there is still room for further improvement. Further work is being done to make the algorithm more adaptive to different types of sequences.

ACKNOWLEDGMENT

This work is supported by the Centre for Multimedia Signal Processing, Hong Kong Polytechnic University and the Research Grant Council of the Hong Kong SAR Government (PolyU 5267/07E).

5. REFERENCES

[1] Ko-Cheung Hui, Wan-Chi Siu and Yui-Lam Chan, “New Adaptive Partial Distortion Search using Clustered Pixel Matching Error Characteristic,” *IEEE Transactions on Image Processing*, vol.14, no.5, pp.597-607, May 2005.

[2] Yui-Lam, Chan and Wan-Chi Siu, “An Efficient Search Strategy for Block Motion Estimation using Image Features,” *IEEE Transactions on Image Processing*, vol.10, no.8, pp.1223-38, Aug. 2001.

[3] Prabhudev Irappa Hosur and Kai-Kuang Ma, “Motion Vector Field Adaptive Fast Motion Estimation,” *Second Int.Conf. on Inf. Comm. and Signal Processing* (ICICS ’99), Dec. 1999.

[4] I. Ahmad, Zheng. Weiguo, Luo. Jiancong, Liou. Ming, “A Fast Adaptive Motion Estimation Algorithm”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol.16, no 3, pp.420-438, Mar. 2006.

[5] Yiu-Lam Chan and Wan-Chi Siu, “New Adaptive Pixel Decimation for Block Motion Vector Estimation,” *IEEE Transactions on Circuits and Systems for Video Technology*, pp.113-118, vol.6, no.1, Feb. 1996.

[6] Ko-Cheung Hui and Wan-Chi Siu, “Extended Analysis of Motion-Compensated Frame Difference for Block-based Motion Prediction Error”, pp.1232-1245, vol.16, No.5, *IEEE Transaction on Image Processing*, May 2007.

[7] Liangming Ji and Wan-Chi Siu, “Reduced Computation using Adaptive Search Window Size for H.264 Multi-frame Motion Estimation”, Paper 1568982117, pp. 1-5, *Proceedings, 14th European Signal Processing conference (EUSIPCO’2006)*, Sep. 2006, Florence Italy

[8] A.C. Yu, K.N. Ngan and G.R. Martin, “Efficient intra- and inter-mode Selection algorithms for H.264/AVC”, J. Vis. Commun. Image R., Vol.17, pp.322-344, 2006.

[9] Libo Yang, Keman Yu, Jiang Li, and Shipeng Li, “An effective variable block-size early termination algorithm for H.264 video coding”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol.15, no.6, June.2005.

[10]S. Saponara, M. Casula, F. Rovati, D. Alfonso, and L. Fanucci, “Dynamic Control of Motion Estimation Search Parameters for Low Complex H.264 Video Coding”, *IEEE Transactions on Consumer Electronics*, vol. 52, no. 1, Feb. 2006.

[11] Ying Zhang, Wan-Chi Siu and Tingzhi Shen, ‘Yet a Faster Motion Estimation Algorithm with Directional Search Strategies’, pp.475-478, Proceedings, 15th Int. Conference on Digital Signal Processing (DSP’2007), July 2007, Cardiff, UK.